整合VLAN與VxLAN之Peregrine軟體定義網路技術

A VLAN-and-VxLAN-based Peregrine SDN Solution

王姿琳
Tzu-Lin Wang

中文摘要

為了讓擁有不同需求的租戶共享實體網路，藉由網路虛擬化技術可虛擬實體網路資源，以提供各個租戶各自的虛擬網路。基於工研院雲端中心研發的初代Peregrine系統(Peregrine 1.0)，本文提出的Peregrine 2.0藉由整合VLAN與VxLAN技術，可使VLAN數量突破先天限制。不同於Peregrine 1.0，本系統加強了動態路徑規劃技術與新增網路流量統計與監控技術、歷史資料查詢等功能。動態路徑規劃進一步考慮實體線路的流量，自動觸發重新規劃與佈建封包路徑，以充分利用實體網路資源。網路流量統計與監控技術則提供網管人員詳細的網路即時資訊，可針對每個租戶、實體線路、資料流或傳輸對的流量進行監控。當網路發生狀況時，網管人員能透過上述資訊快速分析與解決問題。最後，Peregrine 2.0支援高可用性(High Availability)機制，當SDN控制器發生錯誤時，能快速切換至備用控制器，以維持整體網路正常運作。

Abstract

Network virtualization virtualizes a physical network by providing a virtual network with a distinct characteristic to each of the tenants sharing that physical network. This paper describes the design and implementation of a VLAN and VxLAN-based network virtualization system, called Peregrine 2.0, which is based on the VLAN-based Peregrine system (Peregrine 1.0) developed by ITRI CCMA. To eliminate the limitation of VLAN technology, Peregrine 2.0 integrates VxLAN and VLAN technologies to increase the number of VLANs. The Peregrine 2.0 enhances the dynamic traffic engineering mechanism by considering the collected current traffic information, and then triggering the new packet transmission path to be calculated and deployed on the designated network devices. In addition, in Peregrine 2.0 the newly added traffic monitoring mechanism affords the network administrator an unprecedented level of real-time visibility into per-tenant, per-link, and per-flow traffic details, and thus greatly simplifies the root cause analysis of observed network performance problems. Finally, Peregrine 2.0 also supports high availability (HA) such that the network can still work well when SDN controller failed.

關鍵詞(Key Words)

軟體定義網路 (Software-defined Networking；SDN)

網路流量監控 (Network Traffic Monitoring)

高可用性 (High Availability；HA)

虛擬區域網路(Virtual Local Area Network；VLAN)

虛擬區域擴展網路(Virtual Extensible Local Area Network；VxLAN)

1．前言

隨著雲端運算(Cloud Computing)應用的普及，過去的企業網路或傳統資料中心等大型網路環境需要網管人員進行大量網路設定，才能支援整體網路運作。為達此目的，傳統網管人員僅能使用人工設定的方式，逐台連線至網路設備進行配置。然而當網路環境越趨複雜時，網路設定將更加複雜且容易出錯，此現象大幅超過傳統網管人員的負荷；另一方面，為了實現各種網路協定，網路設備需要不斷封裝或解封裝封包，亦導致網路效能不佳。

為了解決上述問題，軟體定義網路(Software-defined Networking；SDN)[1]的虛擬化概念於2008年被提出。SDN採用集中管理的方式各個網路設備，網管人員僅需透過中央控制器(Controller)便能下達指令或封包傳送規則(Rule)至各個網路設備。讓網管人員無需逐台連線至網路設備進行繁瑣的網路設定，藉此可大幅降低人工配置成本與錯誤率。

工研院雲端中心於2014年開發VLAN-based Peregrine軟體定義網路技術[2](以下全文使用”Peregrine 1.0”代稱)並與OpenStack Icehouse[3]進行整合，其主要研發技術為網路虛擬化(Network Virtualization)、動態路徑規劃(Dynamic Traffic Engineering；DTE)及快速故障移轉(Fast Failover；FF)。雖然Peregrine 1.0讓網管人員透過SDN控制器統一下達指令或規則至指定的網路設備，並在網路發生斷路時進行快速故障移轉，以維持網路連通性。但VLAN技術卻存在虛擬網路數量上的限制且不能跨Layer 3網路使用，遠不能滿足大規模資料中心的需求。為了打破上述限制，本論文所提出之新版Peregrine(以下全文使用”Peregrine 2.0”代稱)整合了VLAN與VxLAN技術[4]，VLAN技術可用在網路虛擬化，而VxLAN的標頭(Header)則可擴展虛擬網路數量與支援Layer 3網路的資料傳輸。

此外，為了讓網管人員了解網路即時情況，Peregrine 2.0進一步提供針對每個租戶、虛擬機器、實體與虛擬線路的網路流量監控與統計(Traffic Monitoring；TM)，並將監控結果顯示在Peregrine圖形化使用者介面(Graphical User Interface；GUI)中。當網路壅塞時，網管人員便能透過Peregrine GUI調整封包傳輸路徑，以提升網路整體效能。此外，Peregrine 2.0亦支援高可用性(High Availability)，當SDN控制器發生錯誤時，Peregrine 2.0能快速切換至備用控制器，以維持整體網路運作。

本論文章節架構如下，第2章介紹閱讀本文時需要的相關背景知識；第3章介紹Peregrine 2.0架構與技術細節；第4章介紹Peregrine 2.0使用者介面；最後第5章將總結本文所提出之Peregrine 2.0與其未來發展方向。

2．相關背景知識

本章節分別介紹Peregrine 1.0基本概念與Peregrine 2.0採用的VxLAN技術。

2.1 VLAN-based Peregrine SDN Solution (Peregrine 1.0)

Peregrine 1.0主要開發三種技術，分別為網路虛擬化、動態路徑規劃及快速故障移轉。於網路虛擬化方面，因Peregrine1.0支援OpenStack，每個租戶可自行建立網路(Network)，並對每個網路設定防火牆(Firewall)或閘道器(Gateway)等網路相關配置，並透過VLAN技術隔離每個租戶的網路，使封包傳輸過程中不被干擾。但VLAN技術存在虛擬網路數量的限制且因不能跨Layer 3網路使用，造成資料中心網路配置或擴充困難。於動態路徑規劃方面，Peregrine 1.0的演算法僅根據每條實體

線路(Physical link)之VLAN樹(VLAN Tree)數量進行網路負載平衡計算，其並未真實反應實體拓樸之網路流量，造成所規劃之路徑不一定能使整體網路負載達到平衡。

為了改善Peregrine 1.0不足之處，Peregrine 2.0整合VxLAN與VLAN技術，使虛擬網路數量突破限制；在動態路徑規劃方面，Peregrine 2.0考慮了每條線路的實際流量，當網路壅塞時，Peregrine 2.0將自動更換封包傳輸路徑，以平衡整體網路流量。

2.2 VxLAN技術

VxLAN是一種網路虛擬化技術，用以滿足大型雲端運算環境的擴展性需求，其主要由Cisco、VMware及其他公司共同推出。VxLAN技術將Layer 2網路封包加上VxLAN標頭後，封裝至一個UDP封包中進行傳輸。VxLAN標頭包含一個24位元的ID(VxLAN Network Identifier；VNI)，其作用類似VLAN ID或是GRE Tunnel ID。藉由VNI可輕易突破VLAN的數量限制，足以滿足大規模資料中心的需求。此外，由於VxLAN採用UDP協定封裝，故封包可穿越Layer 3網路，使VxLAN比VLAN擁有更多應用範圍。然而使用VxLAN技術時，終端節點(Endpoint)需對封包進行封裝與解封裝，此舉將使CPU耗費大量時間處理，故若在一個複雜且龐大的網路環境中使用VxLAN技術實現虛擬化，將顯得不切實際。

為了突破VLAN數量限制與VxLAN頻繁封裝或解封裝封包等問題，Peregrine 2.0擷取了VxLAN與VLAN技術的優點，除了使得資料中心可使用的虛擬區域網路數量大幅增加外，亦避免浪費硬體資源，詳細作法將於3.2節中詳細說明。

3 · Peregrine 2.0軟體定義網路技術

本章節分別簡介Peregrine 2.0架構、Peregrine 2.0如何整合VLAN與VxLAN技術、如何實現高可用性與如何進行網路流量統計與監控。

3.1 Peregrine 2.0架構示意圖

圖1為Peregrine 2.0架構圖，其主要使用OpenDaylight Lithium SR2版本[5]作為SDN控制器並與OpenStack Kilo版本[6]整合。網管人員透過OpenStack介面建立虛擬網路與虛擬機器(Virtual Machine；VM)，爾後Peregrine 2.0將觸發其內部模組進行封包路徑規劃，並在Open vSwitch(OvS)與Ethernet交換機設置封包傳輸規則。以下介紹Peregrine 2.0主要流程，其中包含4個重要模組：DTE、FF、TM與DB，分別為動態路徑規劃模組、快速故障轉移模組、網路流量監控模組與資料庫模組。

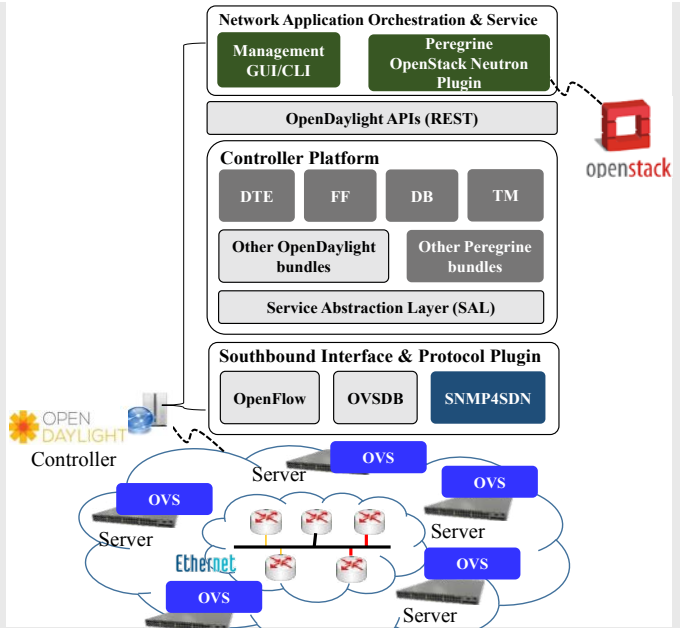


圖 1 Peregrine 2.0架構示意圖

DTE負責計算傳輸路徑並將主要路徑與其備援路徑儲存至DB後，透過OVSDB[7]與工研院雲端中心貢獻的SNMP4SDN[8]專案將路徑規則設定至OvS與乙太網路交換機上。而當線路或交換機發生故障時，FF將自DB取回相對應的備援路徑並將其佈建至網路後，DTE重新計算新的主要路徑與備援路徑存回DB。最後，為了提供詳細的網路流量統計與監控資訊，TM使用OpenFlow[9]設定OvS規

則，若需要監控網路流量時，OvS便將封包鏡射(Mirror)至TM以統計與監控網路流量，詳細作法將於3.4節中詳細說明。

3.2 Peregrine 2.0與VxLAN

為了突破VLAN數量限制與降低VxLAN封包與解封包的影響，Peregrine 2.0擷取兩技術的優點，圖2說明VM傳輸對如何透過Peregrine 2.0進行封包傳輸。圖2中有Zone 1與Zone 2兩個網路區域，兩個Zone各自執行Peregrine 2.0的DTE與FF功能。VLAN-to-VxLAN(V2V)閘道器有VLAN與VxLAN對應表及VM位置資訊，假定目前V2V閘道器1與2內含規則<VxLAN 1，Zone 1，VLAN 100，Zone 2，VLAN 200>，此規則說明Zone 1使用VLAN 100的VM，其傳送封包在Zone 2對應使用VLAN 200進行傳輸。

若Zone 1的VM1預計傳送封包至Zone 2的VM2，則封包傳輸過程如下所述：1)當封包於VM1送出至OvS時，2)OvS將封包的VLAN tag設定為100，並轉送該封包至V2V閘道器1。3)根據規則，V2V閘道器1移除該封包VLAN 100 tag，4)並加上VxLAN 1 ID。當該封包傳送至V2V閘道器2時，5)V2V閘道器2移除封包VxLAN 1 ID，6)並根據規則加上VLAN 200 tag。7)封包透過Ethernet交換機傳送至OvS後，8)OvS移除VLAN 200 tag，並將其傳送至VM2。

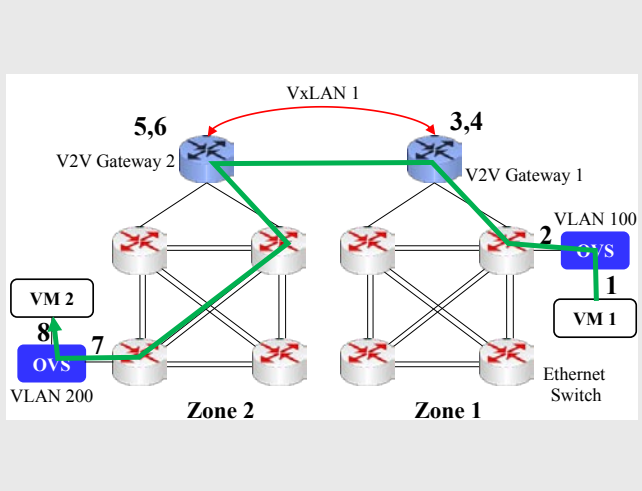


圖 2 Peregrine 2.0 整合VLAN與VxLAN示意圖

藉由上述過程可知，相同的VLAN ID可在不同的Zone中使用，除了突破VLAN數量限制外，由於大部分的VM傳輸對的傳輸並不會跨Zone進行，所以不會有tunnel的負擔。而倘若需要跨Zone傳輸時，Peregrine 2.0亦能保證各網路之間的隔離性。

3.3 Peregrine 2.0與High Availability

當SDN控制器發生錯誤，如何維持網路正常運作是一個非常重要的議題。OpenDaylight於Helium版本開始支援高可用性，當SDN控制器發生錯誤時，可透過Akka技術[10]處理控制器的故障切換。

Peregrine 2.0中有兩個元件使用HA技術，分別為OpenDaylight控制器與MySQL資料庫。圖3為Peregrine 2.0的HA架構示意圖。首先說明Peregrine 2.0控制器的HA元件與架構，本架構以Master-Slave為主。OpenDaylight Cluster Module負責決定哪一台SDN控制器為Master，三台控制器的模組會定期交換訊息，以選出一台Master。若該台Master發生錯誤，其餘的SDN控制器的Cluster模組將另行選出新的Master，而舊有的Master則成為Slave。

Peregrine Master Monitor模組負責監控SDN控制器的角色(Master/Slave)，並根據不同角色作出不同的設定。此程式會不斷詢問本機的Cluster模組以取得本機角色。若本機為Master(如圖中紅線方框所示)，則在本機設定一張擁有虛擬IP的虛擬網卡，Peregrine 2.0可透過此虛擬IP與Master進行連線，而Master上所有與Peregrine2.0相關的模組將被啟動。反之若為Slave，Peregrine Master Monitor模組將刪除本機上具有虛擬IP的虛擬網卡，並關閉Peregrine相關模組，此做法可確保同時間只有一個虛擬IP與一台SDN控制器進行服務。在資料庫方面，Peregrine 2.0使用MySQL Galera Cluster進行資料同步(如圖中藍線所示)，當資料存入Master的MySQL資料庫時，該筆資料亦同步儲存至Slave的MySQL資料庫中。藉由上述機制，

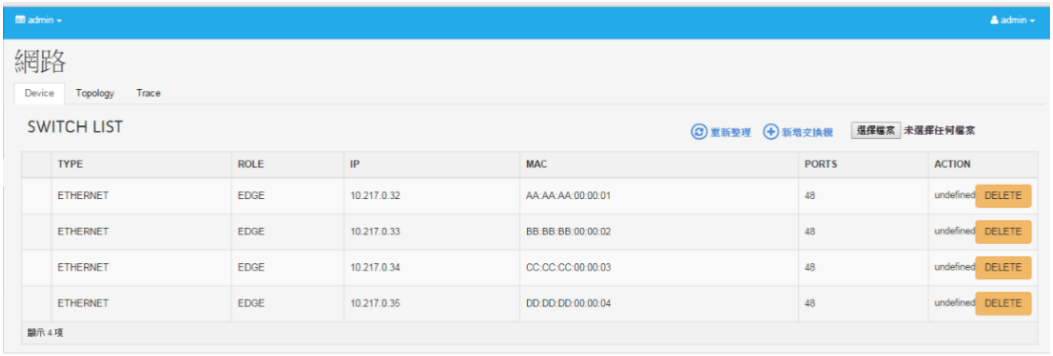


圖 5 網管人員上傳交換機資訊

當控制器發生錯誤時，Peregrine 2.0仍能在切換控制器的過程中保持資料一致性，以維持整體網路運作。

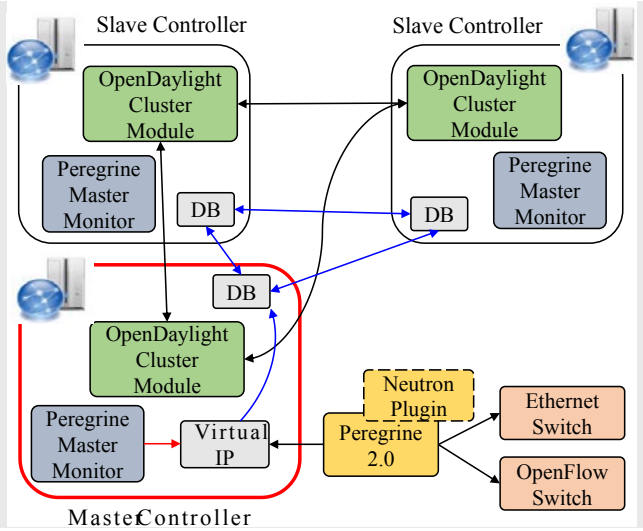


圖 3 Peregrine 2.0 HA架構示意圖

3.4 Peregrine 2.0與Traffic Monitoring

TM模組監控與統計每個虛擬網路、線路與每個VM傳輸對的網路流量，並將結果顯示於Peregrine GUI中。而關於統計與監控流量則分開處理，首先介紹TM監控行為(如圖4藍線所示)，並以監控VM傳輸對的流量為例。當網管人員下指令監控某一VM傳輸對的流量時，該傳輸對的封包會在OvS中被複製一份，被複製的封包其VLAN tag設定為4095。當Monitor Bridge收到此封包後，TM使用tcpdump自Monitor Bridge擷取封包並顯示在GUI上，而原本的封包則持續轉送至目的端(如紅線所示)。另一方面，欲統計某一VM傳輸對流量時，TM透過匹配(Match)OpenFlow規則的封包數量進行流量統計，之後再將結果顯示在GUI上。

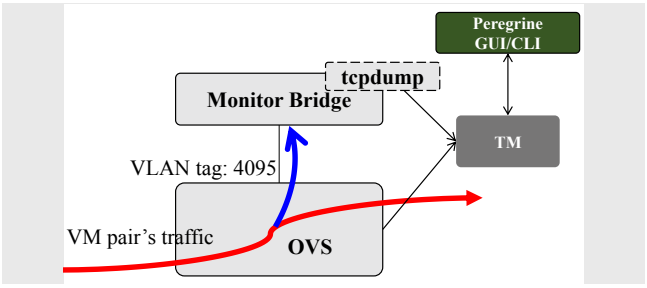


圖 4 Peregrine 2.0 網路流量監控與統計示意圖

為了維持實體線路之負載平衡，DTE可根據每條實體線路的流量，重新規劃封包傳輸路徑。為達此目的，DTE模組允許使用者自訂一個封包流量門檻值。當某一實體線路的流量超過此門檻值時，DTE便自動啟動路徑規劃機制，將大流量實體線路上的虛擬網路其VLAN樹分散至其他流量較低的實體線路上。

4 · Peregrine 2.0使用者介面

如圖5所示，Peregrine 2.0允許網管人員上傳交換機檔案，內容包含交換機型態(Ethernet/OvS)、交換機角色(Core/Edge)、IP、MAC、Port數與使用SNMP需要的相關資訊。除了SNMP相關資訊外，其他資訊都會顯示在GUI上。當網管人員直接在GUI新增或移除交換機後，GUI提供的拓樸資訊亦同時更新。圖6為網路拓樸示意圖，圖中灰線為實體線路，當使用者欲觀察某個VLAN的拓樸時，可藉由GUI選單進行挑選

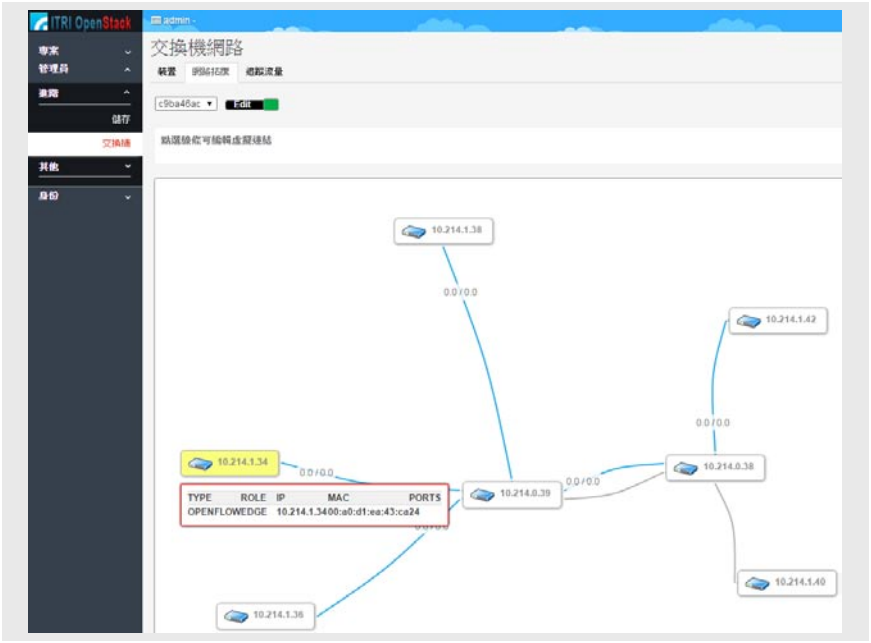


圖 6 顯示VLAN的網路拓樸(如藍線所示)

後，該VLAN的拓樸便能顯示在GUI上(如圖中藍線所示)，而當該VLAN中有封包傳輸時，線路上會顯示目前網路流量。若某一條實體線路的流量過大，流量數字將呈現紅色(如圖7所示)，此時使用者能透過滑鼠選擇合適的灰色實體線路，轉移某一VLAN的傳輸路徑，以維持網路負載平衡。當選好欲替換的傳輸路徑後，網管人員可透過GUI進行DTE的迴圈檢查，若新選的傳輸路徑不會造成迴圈，則GUI通知DTE進行傳輸路徑佈建任務。

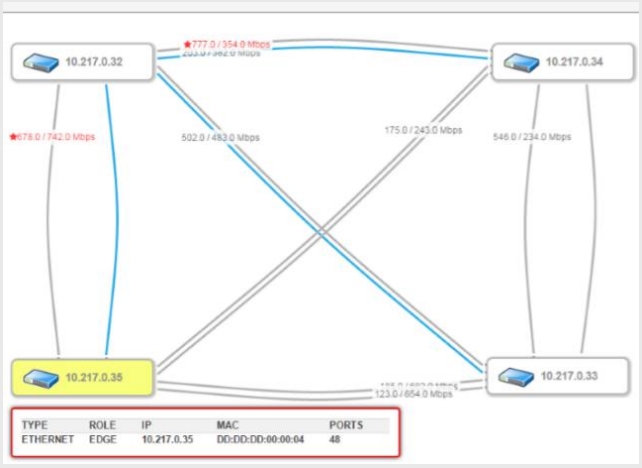


圖 7 網管人員可自訂封包傳輸路徑(如藍線所示)

圖8為流量監控畫面，網管人員能透過此功能得知每個實體線路、虛擬網路、虛擬區域網路與每個VM傳輸對的流量。當網路發生狀況時，網管人員便透過GUI快速查知哪個設備或線路異常，進一步解決問題。

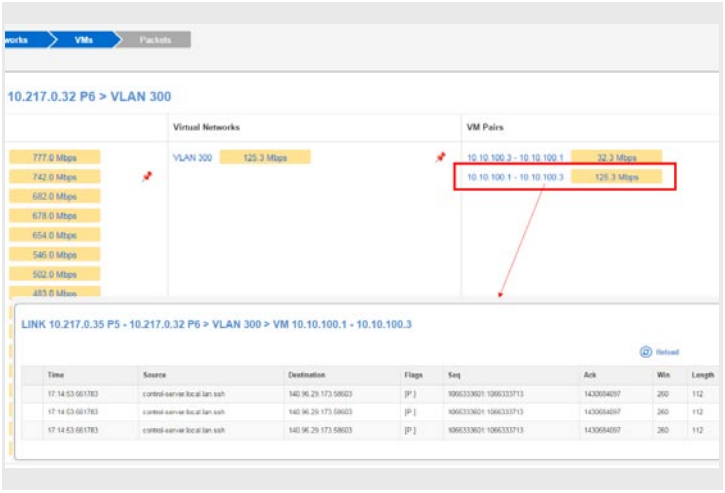


圖 8 流量監控畫面

未來Peregrine 2.0版本的GUI預計加入Peregrine模組健康監控、完整log資訊下載與提供歷史流量與拓樸資訊查詢等功能，以協助網管與Peregrine開發人員進行除錯。

5 · 結論

本文所介紹的Peregrine 2.0是基於工研院所研發之VLAN-based Peregrine系統加以改良，為一個針對傳統乙太網路所開發的一套SDN解決方案。Peregrine 2.0整合VxLAN與VLAN技術並使用Zone的概念，能使其在不增加CPU負擔的情況下，突破VLAN數量限制。另外，Peregrine 2.0新增網路監控與流量功能，使Peregrine 2.0能根據目前網路情況，自動更新封包傳輸路徑。網管人員除了能自由選取欲監控的項目外，亦能根據目前流量手動更換封包傳輸路徑。

未來Peregrine的主要發展方向為加強產品穩定度，開發網管或Peregrine開發人員診斷或除錯工具，以利快速釐清與解決問題。而為了進一產品化，Peregrine 2.0亦目前正積極開發自動化佈建工具，以利網管人員使用。

參考文獻

[1] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, "OpenFlow: Enabling Innovation in Campus Networks," *ACM SIGCOMM Computer Communication Review*, USA, Aug. 2008.

[2] C. Y. Lin, "A VLAN-based Peregrine SDN Solution," *J. Informat. Commun. Technol.*, vol. 161, pp.21-28, 2015.

[3] OpenStack, OpenStack Icehouse [Online]. Available: http://www.openstack.org/software/icehouse

[4] M. Mahalingam, D. Dutt, K. Duda, P. Agarwal, L. Kreeger, T. Sridhar, M. Bursell, and C. Wright, "VXLAN: A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks," Feb. 2012, Internet Draft raft-mahalingam-dutt-dcops-vxlan-01.

[5] OpenDaylight, OpenDaylight Lithium-SR2 [Online]. Available: https://www.opendaylight.org/software/downloads/lithium-sr2

[6] OpenStack, OpenStack Kilo [Online]. Available: https://www.openstack.org/software/kilo

[7] OVSDB, The Open vSwitch Database Management Protocol [Online]. Available: https://tools.ietf.org/html/rfc7047

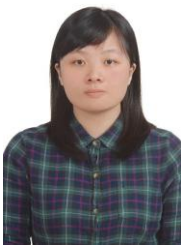
[8] SNMP4SDN, OpenDaylight SNMP4SDN Project [Online]. Available: https://wiki.opendaylight.org/view/SNMP4SDN:Main

[9] OpenFlow, [Online]. Available: https://www.opennetworking.org/index.php

[10] Akka [Online]. Available: http://akka.io/

作者簡介

王姿琳



於2014年取得淡江大學資訊工程學系博士學位，現任工研院資通所資料中心架構與雲端應用軟體組工程師。專長為雲端運算、軟體定義網路及無線感測網路技術。目前從事軟體定義網路技術開發。

E-mail: LinziWang@itri.org.tw